Лекция по эконометрике № 10, часть 1

3 модуль

Tobit models
Sample selection models

Olga Demidova
E-mail:demidova@hse.ru
Department of Applied Economics

План лекции

- Tobit models
- Sample selection models

Tobit Models

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i,$$

$$\Leftrightarrow \quad Y_i^* = x_i'\beta + u_i$$

$$E(u_i) = 0, \ \mathrm{var}(u_i) = \sigma_u^2, \quad i = 1, \ldots, n,$$

$$Y_i = \begin{cases} Y_i^*, \ if \ \ Y_i^* > 0, \\ 0, \ if \ \ Y_i^* \le 0 \end{cases}$$

$$u_i \sim N(0, \sigma_u^2)$$

Tobit Models

$$P(Y_i = 0) = P(Y_i^* \le 0) = P(u_i \le -x_i^{'}\beta) = P(\frac{u_i}{\sigma_u} \le -\frac{x_i^{'}\beta}{\sigma_u}) =$$

$$= \Phi\left(-\frac{x_i^{'}\beta}{\sigma_u}\right) = 1 - \Phi\left(\frac{x_i^{'}\beta}{\sigma_u}\right)$$

Tobit Models

$$The\ distribution\ of\ Y\ truncated\ from\ a\ below$$

$$at\ a\ given\ point\ c:$$

$$f(y \mid Y \geq c) = \frac{f(y)}{P(Y \geq c)}\ if\ y \geq c\ and\ 0\ otherwise.$$

$$If\ Y \sim N(0,1),$$

$$E(Y \mid Y \geq c) = \lambda(c) = \frac{\phi(c)}{1 - \Phi(c)},$$

$$\phi(.), \Phi(.)\ are\ density\ and\ cumulative\ function\ of\ N(0,1).$$

Tobit Models

$$If\ Y \sim N(\mu, \sigma^2),$$

$$E(Y \mid Y > c) = \mu + \sigma\lambda(c^*),\ c^* = \frac{c-\mu}{\sigma}$$

$$For\ tobit\ \text{model}$$

$$E(Y_i \mid Y_i > 0) = x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}$$

Tobit Models

$$For\ tobit\ \mathrm{mod}el$$

$$E(Y_i) = E(Y_i \mid Y_i^* \le 0)P(Y_i^* \le 0) + E(Y_i \mid Y_i^* > 0)P(Y_i^* > 0) =$$

$$= 0 \cdot P(Y_i^* \le 0) + P(u_i > -x_i^{'}\beta)(x_i^{'}\beta + E(u_i \mid u_i > -x_i^{'}\beta)) =$$

$$= \Phi(x_i^{'}\beta / \sigma)x_i^{'}\beta + \sigma\phi(x_i^{'}\beta / \sigma) \ne x_i^{'}\beta$$

Interpretation of Tobit Models

$$E(Y_i) = \Phi(x_i'\beta / \sigma) x_i'\beta + \sigma\phi(x_i'\beta / \sigma) \neq x_i'\beta$$

For tobit mod*el*

$$\frac{\partial E(Y_i)}{\partial X_j} = \Phi(x_i'\beta / \sigma)\beta_j$$

Tobit Models estimation

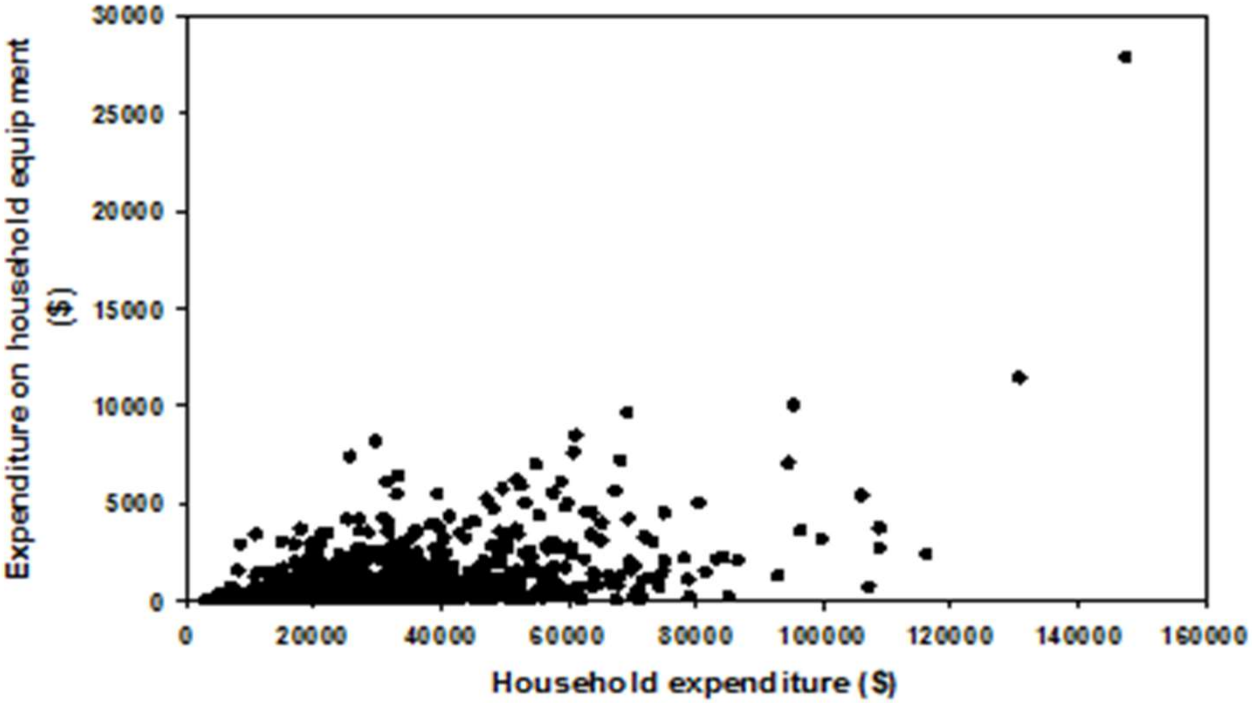
$For\ tobit$ model

$$L(\beta,\sigma^2) = \prod_{Y_i=0} P(Y_i = 0) \prod_{Y_i>0} \phi(Y_i) =$$

$$= \prod_{Y_i=0} \left(1 - \Phi\left(\frac{x_i^{'}\beta}{\sigma_u}\right)\right) \prod_{Y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_i - x_i^{'}\beta)^2\right\} \to \max$$

Tobit Models. Example (Copyright Christopher Dougherty 2001–2006)

TOBIT ANALYSIS



We will use the Consumer Expenditure Survey data set to illustrate the use of tobit analysis. The figure plots annual household expenditure on household equipment, *HEQ*, on total household expenditure, *EXP*, both measured in dollars.

31

Tobit Models. Example

TOBIT ANALYSIS

```
. tab HEQ if HEQ<10
```

HEQ	Freq.	Percent	Cum.
0	86	89.58	89.58
3	1	1.04	90.62
4	2	2.08	92.71
6	1	1.04	93.75
7	1	1.04	94.79
8	5	5.21	100.00
Total	96	100.00	

For 86 households, *HEQ* was 0. (The tabulation has been confined to small values of *HEQ*. We are only interested in finding out how many actually had *HEQ* = 0.)

32

Tobit Models. Example

TOBIT ANALYSIS

```
. reg HEQ EXP

  Source |       SS       df       MS              Number of obs =     869
---------+------------------------------           F(  1,   867) =  353.91
   Model |   729289164     1   729289164           Prob > F      =  0.0000
Residual |  1.7866e+09   867  2060635.12           R-squared     =  0.2899
---------+------------------------------           Adj R-squared =  0.2891
   Total |  2.5159e+09   868  2898456.01           Root MSE      =  1435.5

------------------------------------------------------------------------------
     HEQ |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     EXP |   .0471546   .0025065     18.813   0.000        .042235    .0520742
   _cons |  -397.2088   89.44449     -4.441   0.000      -572.7619   -221.6558
------------------------------------------------------------------------------
```

Here is a regression using all the observations. We anticipate that the coefficient of *EXP* is biased downwards.

33

Tobit Models. Example

TOBIT ANALYSIS

```
. reg HEQ EXP if HEQ>0

  Source |       SS       df       MS              Number of obs =     783
---------+------------------------------           F(  1,   781) =  291.04
   Model |  656349265      1  656349265            Prob > F      =  0.0000
Residual |  1.7613e+09   781  2255219.19           R-squared     =  0.2715
---------+------------------------------           Adj R-squared =  0.2705
   Total |  2.4177e+09   782  3091656.59           Root MSE      =  1501.7

------------------------------------------------------------------------------
     HEQ |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     EXP |   .0467672   .0027414     17.060   0.000       .0413859    .0521485
   _cons |  -350.1704   101.8034     -3.440   0.001      -550.0112   -150.3296
------------------------------------------------------------------------------
```

Here is an OLS regression with the constrained observations dropped. The estimate of the slope coefficient is almost the same, just a little lower.

34

Tobit Models. Example

TOBIT ANALYSIS

```
. tobit HEQ EXP, ll(0)

Tobit Estimates                                   Number of obs =       869
                                                  chi2(1)       =    315.41
                                                  Prob > chi2   =    0.0000
Log Likelihood = -6911.0175                       Pseudo R2     =    0.0223

------------------------------------------------------------------------------
     HEQ |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     EXP |   .0520828   .0027023     19.273   0.000       .0467789    .0573866
   _cons |  -661.8156   97.95977     -6.756   0.000      -854.0813   -469.5499
---------+--------------------------------------------------------------------
     _se |   1521.896    38.6333              (Ancillary parameter)
------------------------------------------------------------------------------

Obs. summary:        86  left-censored observations at HEQ<=0
                    783     uncensored observations
```

Here is the tobit regression. The Stata command is 'tobit', followed by the dependent variable and the explanatory variables, then a comma, then 'll' and in parentheses the lower limit.

35

Tobit Models. Example

TOBIT ANALYSIS

```
. tobit HEQ EXP, ll(0)

Tobit Estimates                                   Number of obs =     869
                                                  chi2(1)       =  315.41
                                                  Prob > chi2   =  0.0000
Log Likelihood = -6911.0175                       Pseudo R2     =  0.0223

------------------------------------------------------------------------------
     HEQ |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     EXP |   .0520828   .0027023     19.273   0.000       .0467789    .0573866
   _cons |  -661.8156   97.95977     -6.756   0.000      -854.0813   -469.5499
---------+--------------------------------------------------------------------
     _se |   1521.896    38.6333              (Ancillary parameter)
------------------------------------------------------------------------------

Obs. summary:        86 left-censored observations at HEQ<=0
                    783 uncensored observations
```

If the dependent variable were constrained by an upper limit, we would use 'ul' instead of 'll', with the upper limit in parentheses. The method can handle lower limits and upper limits simultaneously.

36

Tobit Models. Example

TOBIT ANALYSIS

```
. tobit HEQ EXP, ll(0)
------------------------------------------------------------------------------
         HEQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         EXP |   .0520828   .0027023    19.273   0.000     .0467789    .0573866
       _cons |  -661.8156   97.95977    -6.756   0.000    -854.0813   -469.5499
-------------+----------------------------------------------------------------
         _se |   1521.896    38.6333             (Ancillary parameter)
------------------------------------------------------------------------------
. reg HEQ EXP
------------------------------------------------------------------------------
         HEQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         EXP |   .0471546   .0025065    18.813   0.000      .042235    .0520742
       _cons |  -397.2088   89.44449    -4.441   0.000    -572.7619   -221.6558
------------------------------------------------------------------------------
. reg HEQ EXP if HEQ>0
------------------------------------------------------------------------------
         HEQ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         EXP |   .0467672   .0027414    17.060   0.000     .0413859    .0521485
       _cons |  -350.1704   101.8034    -3.440   0.001    -550.0112   -150.3296
------------------------------------------------------------------------------
```

We see that the coefficient of *EXP* is indeed larger in the tobit analysis, confirming the downwards bias in the OLS estimates. In this case the difference is not very great. That is because only 10 percent of the observations were constrained.

37

Sample selection model

$$Y_i^* = x_i^{'}\beta + \varepsilon_i$$

$$g_i^* = z_i^{'}\gamma + u_i$$

$$g_i = \begin{cases} 1, \ if \ \ g_i^* \geq 0 \\ 0, \ if \ \ g_i^* < 0 \end{cases},$$

$$General \quad \mathrm{mod}\, el:$$

$$\begin{cases} Y_i = Y_i^*, g = 1, if \ \ g_i^* \geq 0 \\ Y_i \ is \ not \ observed \ , g = 0, if \ \ g_i^* < 0 \end{cases},$$

$$Let \ \ (\varepsilon_i, u_i) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & 1 \end{pmatrix}\right).$$

Sample selection model

$$E(Y_i \mid g_i = 1) = x_i^{'}\beta + E(\varepsilon_i \mid g_i = 1) =$$

$$= x_i^{'}\beta + E(\varepsilon_i \mid u_i \geq -z_i^{'}\gamma),$$

$$Lemma\ .\ If\ \ \varepsilon_i = \sigma_{\varepsilon u} u_i + \eta_i$$

$$then\ \ u_i\ \ and\ \ \eta_i\ \ are\ \ independan\ \ t.$$

$$\Pr\ oof\ . \triangleright \operatorname{cov}(u_i, \eta_i) = Eu_i \eta_i = E(u_i(\varepsilon_i - \sigma_{\varepsilon u} u_i)) =$$

$$= Eu_i \varepsilon_i - \sigma_{\varepsilon u} E(u_i^2) = \sigma_{\varepsilon u} - \sigma_{\varepsilon u} \sigma_u^2 = 0 \triangleleft .$$

$$E(Y_i \mid g_i = 1) = x_i^{'}\beta + E(\sigma_{\varepsilon u} u_i + \eta_i \mid u_i \geq -z_i^{'}\gamma) =$$

$$= x_i^{'}\beta + \sigma_{\varepsilon u} E(u_i \mid u_i \geq -z_i^{'}\gamma) = x_i^{'}\beta + \sigma_{\varepsilon u} \lambda(z_i^{'}\gamma)$$

$$where\ \ \lambda(z_i^{'}\gamma) = \frac{\phi(z_i^{'}\gamma)}{\Phi(z_i^{'}\gamma)}\ is\ \ Heckman\ 's\ lambda\ .$$

Sample selection model. Estimation

1) ML

$$l(\beta,\gamma,\sigma_\varepsilon^2,\sigma_{\varepsilon u}) = \sum_{g_i=0} \ln P(g_i=0) + \sum_{g_i=1} \ln f(Y_i, g_i = 1) =$$

$$= \sum_{g_i=0} \ln P(g_i=0) + \sum_{g_i=1} \ln[P(g_i=1 \mid Y_i) f(Y_i)] =$$

$$= \sum_{g_i=0} \ln P(g_i=0) + \sum_{g_i=1} [\ln P(g_i=1 \mid Y_i) + \ln f(Y_i)] \to \max$$

$$P(g_i=0) = 1 - \Phi(z_i'\gamma),$$

$$f(Y_i) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left\{-\frac{(Y_i - x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}$$

$$P(g_i=1 \mid Y_i) = \Phi\left(\frac{z_i'\gamma + (\sigma_{\varepsilon u}/\sigma_\varepsilon^2)(Y_i - x_i'\beta)}{\sqrt{1-\sigma_{\varepsilon u}^2/\sigma_\varepsilon^2}}\right)$$

Sample selection model. Estimation

$$2)\quad Two - step \quad procedure$$

$$1st \quad step \qquad g_i^* = z_i^{'}\gamma + u_i$$

$$g_i = \begin{cases} 1, if \quad g_i^* \geq 0 \\ 0, if \quad g_i^* < 0 \end{cases},$$

$$2nd \quad step \quad :$$

$$Y_i = x_i^{'}\beta + \sigma_{\varepsilon u}\lambda(z_i^{'}\hat{\gamma}) + \xi_i$$

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

We will illustrate the heckman procedure by fitting an earnings function for females using the LFP data set on the website. The includes 2,661 females, of whom 2,021 had earnings, in 1994.

19

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

In Stata the regression command is 'heckman'.

20

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

After a comma, the selection process is specified using 'select' followed by the selection variables in parentheses.

22

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

CHILDL06 is a dummy equal to 1 if there is a child aged less than 6. *CHILDL16* is 1 if there is a child aged less than 15 and *CHILDL06* is 0. *MARRIED* is equal to 1 if the respondent is married with spouse present. Otherwise the variables are as defined in the *EAEF* data sets.

23

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                         Number of obs      =      2661
(regression model with sample selection)        Censored obs       =       640
                                                Uncensored obs     =      2021
                                                Wald chi2(4)       =    714.73
Log likelihood =  -2668.81                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
LGEARN   |
       S |    .095949    .0056438    17.001   0.000       .0848874    .1070106
  ASVABC |   .0110391    .0014658     7.531   0.000       .0081663    .0139119
ETHBLACK |   -.066425    .0381626    -1.741   0.082      -.1412223    .0083722
 ETHHISP |   .0744607    .0450095     1.654   0.098      -.0137563    .1626777
   _cons |   4.901626    .0768254    63.802   0.000       4.751051    5.052202
---------+--------------------------------------------------------------------
select   |
       S |   .1041415    .0119836     8.690   0.000       .0806541    .1276288
     AGE |  -.0357225     .011105    -3.217   0.001      -.0574879   -.0139572
CHILDL06 |  -.3982738    .0703418    -5.662   0.000      -.5361412   -.2604064
CHILDL16 |   .0254818    .0709693     0.359   0.720      -.1136155     .164579
 MARRIED |   .0121171    .0546561     0.222   0.825      -.0950069    .1192412
ETHBLACK |  -.2941378    .0787339    -3.736   0.000      -.4484535   -.1398222
 ETHHISP |  -.0178776    .1034237    -0.173   0.863      -.2205843    .1848292
   _cons |   .1682515    .2606523     0.646   0.519      -.3426176    .6791206
---------+--------------------------------------------------------------------
```

The numbers of participating and non-participating respondents are given at the top of the output.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                         Number of obs      =      2661
(regression model with sample selection)        Censored obs       =       640
                                                Uncensored obs     =      2021
                                                Wald chi2(4)       =    714.73
Log likelihood =  -2668.81                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
LGEARN       |
           S |    .095949   .0056438    17.001   0.000      .0848874    .1070106
      ASVABC |   .0110391   .0014658     7.531   0.000      .0081663    .0139119
    ETHBLACK |   -.066425   .0381626    -1.741   0.082     -.1412223    .0083722
     ETHHISP |   .0744607   .0450095     1.654   0.098     -.0137563    .1626777
       _cons |   4.901626   .0768254    63.802   0.000      4.751051    5.052202
-------------+----------------------------------------------------------------
select       |
           S |   .1041415   .0119836     8.690   0.000      .0806541    .1276288
         AGE |  -.0357225    .011105    -3.217   0.001     -.0574879   -.0139572
    CHILDL06 |  -.3982738   .0703418    -5.662   0.000     -.5361412   -.2604064
    CHILDL16 |   .0254818   .0709693     0.359   0.720     -.1136155     .164579
     MARRIED |   .0121171   .0546561     0.222   0.825     -.0950069    .1192412
    ETHBLACK |  -.2941378   .0787339    -3.736   0.000     -.4484535   -.1398222
     ETHHISP |  -.0178776   .1034237    -0.173   0.863     -.2205843    .1848292
       _cons |   .1682515   .2606523     0.646   0.519     -.3426176    .6791206
-------------+----------------------------------------------------------------
```

Next comes the regression output.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                          Number of obs      =      2661
(regression model with sample selection)         Censored obs       =       640
                                                 Uncensored obs     =      2021
                                                 Wald chi2(4)       =    714.73
Log likelihood =  -2668.81                       Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
LGEARN   |
       S |    .095949   .0056438    17.001   0.000      .0848874    .1070106
  ASVABC |   .0110391   .0014658     7.531   0.000      .0081663    .0139119
ETHBLACK |   -.066425   .0381626    -1.741   0.082     -.1412223    .0083722
 ETHHISP |   .0744607   .0450095     1.654   0.098     -.0137563    .1626777
   _cons |   4.901626   .0768254    63.802   0.000      4.751051    5.052202
---------+--------------------------------------------------------------------
select   |
       S |   .1041415   .0119836     8.690   0.000      .0806541    .1276288
     AGE |  -.0357225    .011105    -3.217   0.001     -.0574879   -.0139572
CHILDL06 |  -.3982738   .0703418    -5.662   0.000     -.5361412   -.2604064
CHILDL16 |   .0254818   .0709693     0.359   0.720     -.1136155     .164579
 MARRIED |   .0121171   .0546561     0.222   0.825     -.0950069    .1192412
ETHBLACK |  -.2941378   .0787339    -3.736   0.000     -.4484535   -.1398222
 ETHHISP |  -.0178776   .1034237    -0.173   0.863     -.2205843    .1848292
   _cons |   .1682515   .2606523     0.646   0.519     -.3426176    .6791206
---------+--------------------------------------------------------------------
```

The results of the probit analysis of the selection process follow.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                          Number of obs      =      2661
           |      Coef.   Std. Err.        z     P>|z|      [95% Conf. Interval]
-----------+----------------------------------------------------------------------
select     |
         S |   .1041415   .0119836      8.690   0.000      .0806541    .1276288
       AGE |  -.0357225    .011105     -3.217   0.001     -.0574879   -.0139572
  CHILDL06 |  -.3982738   .0703418     -5.662   0.000     -.5361412   -.2604064
  CHILDL16 |   .0254818   .0709693      0.359   0.720     -.1136155     .164579
   MARRIED |   .0121171   .0546561      0.222   0.825     -.0950069    .1192412
  ETHBLACK |  -.2941378   .0787339     -3.736   0.000     -.4484535   -.1398222
   ETHHISP |  -.0178776   .1034237     -0.173   0.863     -.2205843    .1848292
     _cons |   .1682515   .2606523      0.646   0.519     -.3426176    .6791206
-----------+----------------------------------------------------------------------
   /athrho |    1.01804   .0932533     10.917   0.000      .8352669    1.200813
  /lnsigma |  -.6349788   .0247858    -25.619   0.000     -.6835582   -.5863994
-----------+----------------------------------------------------------------------
       rho |    .769067   .0380973                          .683294    .8339024
     sigma |   .5299467   .0131352                         .5048176    .5563268
    lambda |   .4075645     .02867                         .3513724    .4637567
----------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    32.90   Prob > chi2 = 0.0000
----------------------------------------------------------------------------------
```

The final part of the output gives the selection bias statistics. rho gives an estimate of the correlation between *ε* and *u*, here 0.77.

28

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                       Number of obs      =      2661
             |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
select       |
           S |   .1041415   .0119836     8.690   0.000      .0806541   .1276288
         AGE |  -.0357225    .011105    -3.217   0.001     -.0574879  -.0139572
     CHILDL06 |  -.3982738   .0703418    -5.662   0.000     -.5361412  -.2604064
     CHILDL16 |   .0254818   .0709693     0.359   0.720     -.1136155    .164579
     MARRIED |   .0121171   .0546561     0.222   0.825     -.0950069   .1192412
    ETHBLACK |  -.2941378   .0787339    -3.736   0.000     -.4484535  -.1398222
     ETHHISP |  -.0178776   .1034237    -0.173   0.863     -.2205843   .1848292
       _cons |   .1682515   .2606523     0.646   0.519     -.3426176   .6791206
-------------+----------------------------------------------------------------
     /athrho |    1.01804   .0932533    10.917   0.000      .8352669   1.200813
    /lnsigma |  -.6349788   .0247858   -25.619   0.000     -.6835582  -.5863994
-------------+----------------------------------------------------------------
         rho |    .769067   .0380973                        .683294   .8339024
       sigma |   .5299467   .0131352                       .5048176   .5563268
      lambda |   .4075645     .02867                       .3513724   .4637567
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    32.90   Prob > chi2 = 0.0000
------------------------------------------------------------------------------
```

For technical reasons, ρ is estimated indirectly via atanh ρ. However, a test of the null hypothesis H_0: atanh $\rho = 0$ is equivalent to a test of the null hypothesis of H_0: $\rho = 0$.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

Heckman selection model — Number of obs = 2661

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
select						
S	.1041415	.0119836	8.690	0.000	.0806541	.1276288
AGE	-.0357225	.011105	-3.217	0.001	-.0574879	-.0139572
CHILDL06	-.3982738	.0703418	-5.662	0.000	-.5361412	-.2604064
CHILDL16	.0254818	.0709693	0.359	0.720	-.1136155	.164579
MARRIED	.0121171	.0546561	0.222	0.825	-.0950069	.1192412
ETHBLACK	-.2941378	.0787339	-3.736	0.000	-.4484535	-.1398222
ETHHISP	-.0178776	.1034237	-0.173	0.863	-.2205843	.1848292
_cons	.1682515	.2606523	0.646	0.519	-.3426176	.6791206
/athrho	1.01804	.0932533	10.917	0.000	.8352669	1.200813
/lnsigma	-.6349788	.0247858	-25.619	0.000	-.6835582	-.5863994
rho	.769067	.0380973			.683294	.8339024
sigma	.5299467	.0131352			.5048176	.5563268
lambda	.4075645	.02867			.3513724	.4637567

LR test of indep. eqns. (rho = 0): chi2(1) = 32.90 Prob > chi2 = 0.0000

The asymptotic *t* statistic is 10.92 and so the null hypothesis is rejected.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                          Number of obs      =      2661
(regression model with sample selection)         Censored obs       =       640
                                                 Uncensored obs     =      2021
                                                 Wald chi2(4)       =    714.73
Log likelihood =  -2668.81                       Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
LGEARN   |
       S |    .095949   .0056438    17.001   0.000      .0848874    .1070106
  ASVABC |   .0110391   .0014658     7.531   0.000      .0081663    .0139119
ETHBLACK |   -.066425   .0381626    -1.741   0.082     -.1412223    .0083722
 ETHHISP |   .0744607   .0450095     1.654   0.098     -.0137563    .1626777
   _cons |   4.901626   .0768254    63.802   0.000      4.751051    5.052202
---------+--------------------------------------------------------------------
select   |
       S |   .1041415   .0119836     8.690   0.000      .0806541    .1276288
     AGE |  -.0357225    .011105    -3.217   0.001     -.0574879   -.0139572
CHILDL06 |  -.3982738   .0703418    -5.662   0.000     -.5361412   -.2604064
CHILDL16 |   .0254818   .0709693     0.359   0.720     -.1136155     .164579
 MARRIED |   .0121171   .0546561     0.222   0.825     -.0950069    .1192412
ETHBLACK |  -.2941378   .0787339    -3.736   0.000     -.4484535   -.1398222
 ETHHISP |  -.0178776   .1034237    -0.173   0.863     -.2205843    .1848292
   _cons |   .1682515   .2606523     0.646   0.519     -.3426176    .6791206
---------+--------------------------------------------------------------------
```

An alternative test involves a comparison of the log likelihood for this model with that for a restricted version where ρ is assumed to be 0.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

Heckman selection model — Number of obs = 2661

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
select						
S	.1041415	.0119836	8.690	0.000	.0806541	.1276288
AGE	-.0357225	.011105	-3.217	0.001	-.0574879	-.0139572
CHILDL06	-.3982738	.0703418	-5.662	0.000	-.5361412	-.2604064
CHILDL16	.0254818	.0709693	0.359	0.720	-.1136155	.164579
MARRIED	.0121171	.0546561	0.222	0.825	-.0950069	.1192412
ETHBLACK	-.2941378	.0787339	-3.736	0.000	-.4484535	-.1398222
ETHHISP	-.0178776	.1034237	-0.173	0.863	-.2205843	.1848292
_cons	.1682515	.2606523	0.646	0.519	-.3426176	.6791206
/athrho	1.01804	.0932533	10.917	0.000	.8352669	1.200813
/lnsigma	-.6349788	.0247858	-25.619	0.000	-.6835582	-.5863994
rho	.769067	.0380973			.683294	.8339024
sigma	.5299467	.0131352			.5048176	.5563268
lambda	.4075645	.02867			.3513724	.4637567

LR test of indep. eqns. (rho = 0): chi2(1) = 32.90 Prob > chi2 = 0.0000

The test statistic 2 (log L_U – log L_R), where log L_U and log L_R are the log-likelihoods for the unrestricted and restricted versions, is distributed as a chi-squared statistic with 1 degree of freedom under the null hypothesis that the restriction $\rho = 0$ is valid.

32

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

Heckman selection model — Number of obs = 2661

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
select						
S	.1041415	.0119836	8.690	0.000	.0806541	.1276288
AGE	-.0357225	.011105	-3.217	0.001	-.0574879	-.0139572
CHILDL06	-.3982738	.0703418	-5.662	0.000	-.5361412	-.2604064
CHILDL16	.0254818	.0709693	0.359	0.720	-.1136155	.164579
MARRIED	.0121171	.0546561	0.222	0.825	-.0950069	.1192412
ETHBLACK	-.2941378	.0787339	-3.736	0.000	-.4484535	-.1398222
ETHHISP	-.0178776	.1034237	-0.173	0.863	-.2205843	.1848292
_cons	.1682515	.2606523	0.646	0.519	-.3426176	.6791206
/athrho	1.01804	.0932533	10.917	0.000	.8352669	1.200813
/lnsigma	-.6349788	.0247858	-25.619	0.000	-.6835582	-.5863994
rho	.769067	.0380973			.683294	.8339024
sigma	.5299467	.0131352			.5048176	.5563268
lambda	.4075645	.02867			.3513724	.4637567

LR test of indep. eqns. (rho = 0): chi2(1) = 32.90 Prob > chi2 = 0.0000

In this example the test statistic is 32.90. The critical value of chi-squared with one degree of freedom at the 0.1 percent level is 10.83, so the null hypothesis is rejected.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
------------------------------------------------------------------------------
             |      Coef.   Std. Err.        z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
LGEARN       |
           S |    .095949   .0056438    17.001   0.000      .0848874    .1070106
      ASVABC |   .0110391   .0014658     7.531   0.000      .0081663    .0139119
    ETHBLACK |   -.066425   .0381626    -1.741   0.082     -.1412223    .0083722
     ETHHISP |   .0744607   .0450095     1.654   0.098     -.0137563    .1626777
       _cons |   4.901626   .0768254    63.802   0.000      4.751051    5.052202
---------+--------------------------------------------------------------------

. reg LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0
------------------------------------------------------------------------------
      LGEARN |      Coef.   Std. Err.        t     P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
           S |   .0807836    .005244    15.405   0.000      .0704994    .0910677
      ASVABC |   .0117377   .0014886     7.885   0.000      .0088184     .014657
    ETHBLACK |  -.0148782   .0356868    -0.417   0.677     -.0848649    .0551086
     ETHHISP |   .0802266    .041333     1.941   0.052     -.0008333    .1612865
       _cons |   5.223712   .0703534    74.250   0.000      5.085739    5.361685
------------------------------------------------------------------------------
```

It is instructive to compare the fitted earnings functions for the heckman and least squares models. The coefficients are fairly similar, despite the inconsistency of the least squares estimates.

34

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LGEARN						
S	.095949	.0056438	17.001	0.000	.0848874	.1070106
ASVABC	.0110391	.0014658	7.531	0.000	.0081663	.0139119
ETHBLACK	-.066425	.0381626	-1.741	0.082	-.1412223	.0083722
ETHHISP	.0744607	.0450095	1.654	0.098	-.0137563	.1626777
_cons	4.901626	.0768254	63.802	0.000	4.751051	5.052202

```
. reg LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0
```

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.0807836	.005244	15.405	0.000	.0704994	.0910677
ASVABC	.0117377	.0014886	7.885	0.000	.0088184	.014657
ETHBLACK	-.0148782	.0356868	-0.417	0.677	-.0848649	.0551086
ETHHISP	.0802266	.041333	1.941	0.052	-.0008333	.1612865
_cons	5.223712	.0703534	74.250	0.000	5.085739	5.361685

The coefficient of schooling is a little higher in the heckman regression.

35

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
Heckman selection model                       Number of obs      =      2661
(regression model with sample selection)      Censored obs       =       640
                                              Uncensored obs     =      2021
                                              Wald chi2(4)       =    714.73
Log likelihood =  -2668.81                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
LGEARN   |
       S |    .095949   .0056438   17.001   0.000      .0848874    .1070106
  ASVABC |   .0110391   .0014658    7.531   0.000      .0081663    .0139119
ETHBLACK |   -.066425   .0381626   -1.741   0.082     -.1412223    .0083722
 ETHHISP |   .0744607   .0450095    1.654   0.098     -.0137563    .1626777
   _cons |   4.901626   .0768254   63.802   0.000      4.751051    5.052202
---------+--------------------------------------------------------------------
select   |
       S |   .1041415   .0119836    8.690   0.000      .0806541    .1276288
     AGE |  -.0357225    .011105   -3.217   0.001     -.0574879   -.0139572
 CHILDL06|  -.3982738   .0703418   -5.662   0.000     -.5361412   -.2604064
 CHILDL16|   .0254818   .0709693    0.359   0.720     -.1136155     .164579
  MARRIED|   .0121171   .0546561    0.222   0.825     -.0950069    .1192412
ETHBLACK |  -.2941378   .0787339   -3.736   0.000     -.4484535   -.1398222
 ETHHISP |  -.0178776   .1034237   -0.173   0.863     -.2205843    .1848292
   _cons |   .1682515   .2606523    0.646   0.519     -.3426176    .6791206
---------+--------------------------------------------------------------------
```

The probit analysis showed that schooling has a highly significant positive effect on labor force participation, controlling for other characteristics such as number of children of school age.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
LGEARN       |
           S |    .095949   .0056438   17.001   0.000      .0848874    .1070106
      ASVABC |   .0110391   .0014658    7.531   0.000      .0081663    .0139119
    ETHBLACK |   -.066425   .0381626   -1.741   0.082     -.1412223    .0083722
     ETHHISP |   .0744607   .0450095    1.654   0.098     -.0137563    .1626777
       _cons |   4.901626   .0768254   63.802   0.000      4.751051    5.052202
-------------+----------------------------------------------------------------

. reg LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0
------------------------------------------------------------------------------
      LGEARN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   .0807836    .005244   15.405   0.000      .0704994    .0910677
      ASVABC |   .0117377   .0014886    7.885   0.000      .0088184     .014657
    ETHBLACK |  -.0148782   .0356868   -0.417   0.677     -.0848649    .0551086
     ETHHISP |   .0802266    .041333    1.941   0.052     -.0008333    .1612865
       _cons |   5.223712   .0703534   74.250   0.000      5.085739    5.361685
------------------------------------------------------------------------------
```

If females with higher levels of schooling are relatively keen to work, they will tend to be willing to accept lower wages, controlling for other factors including education, than those who are reluctant to work.

Sample selection model. Example (C.Dougherty)

SAMPLE SELECTION BIAS

```
. heckman LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0, select(S AGE CHILDL06
CHILDL16 MARRIED ETHBLACK ETHHISP)
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LGEARN						
S	.095949	.0056438	17.001	0.000	.0848874	.1070106
ASVABC	.0110391	.0014658	7.531	0.000	.0081663	.0139119
ETHBLACK	-.066425	.0381626	-1.741	0.082	-.1412223	.0083722
ETHHISP	.0744607	.0450095	1.654	0.098	-.0137563	.1626777
_cons	4.901626	.0768254	63.802	0.000	4.751051	5.052202

```
. reg LGEARN S ASVABC ETHBLACK ETHHISP if MALE==0
```

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.0807836	.005244	15.405	0.000	.0704994	.0910677
ASVABC	.0117377	.0014886	7.885	0.000	.0088184	.014657
ETHBLACK	-.0148782	.0356868	-0.417	0.677	-.0848649	.0551086
ETHHISP	.0802266	.041333	1.941	0.052	-.0008333	.1612865
_cons	5.223712	.0703534	74.250	0.000	5.085739	5.361685

Hence the wages of more-educated females will tend not to reflect the full value of education in the market place. The least squares regression does not take account of this, and hence the estimate of the return to schooling is lower.